Editorial decision

Dear authors,

I have reached a decision concerning your preprint number 457, entitled "The contribution of Mediterranean connectivity to morphological variability in Iron Age sheep of the Eastern Mediterranean" after receiving the reviewers comments.

Although all reviewers agreed on the interest and quality of the work presented in your manuscript, they feel the need to ask for specific points to be addressed before acceptance. I suggest you carefully address these and resubmit a modified version of the manuscript.

You should find reviewers' comments attached to the present email.

Looking forward receiving your modified version,

Kind regards,

Louise Le Meillour for PCI Archaeology

*by **Louise Le Meillour**, 06 Feb 2024 11:45*
Manuscript: **https://doi.org/10.1101/2022.12.24.521859**
version: 2

Reviewer #1

Overall, the paper's format needs revision. There are fragments of the introduction found in the materials & methods section, and vice versa. This mixing of sections makes it challenging for readers to follow.

Thank you for these comments, which we accepted as detailed below.

Additionally, the statistical method is not well-explained, making it difficult to comprehend the obtained results. The results are inadequately discussed in the conclusion; incorporating comparisons with bibliographic references could add weight to the findings. These various problems of form make it difficult to appreciate the substance.

We added references and explanations. See below in the section-specific comments.

The sample sizes for the analyses conducted are adequate. However, it is regrettable that other sites, particularly from Cyprus, are not included (if possible). Including additional sites could have strengthened the obtained results and helped account for the potential bias of the island effect.

We agree. While a larger number of Iron Age sites with numerous sheep astragali would have improved our study, it was beyond us in terms of archaeological material availability and funding.

The discussion does not clearly address the question regarding variance. It would be beneficial to enhance clarity in both the stated objectives and, more importantly, in the provided answers.

This is not very clear, but the discussion has been revised extensively, in a way that we hope answers this comment.

Abstract & Introduction
The abstract lacks a clear link between the scarcity of studies on Mediterranean fauna and maritime connections in the Eastern Mediterranean.

The connection was emphasized by rephrasing the first two sentences of the abstract, which now read:

“The movement of livestock across the Mediterranean is well-documented in the Neolithic era, but its significance during subsequent periods has received less attention. Here we start addressing this lacuna by analyzing astragal bone morphology from four coastal and inland sites in Israel and Cyprus, seeking potential evidence for maritime connections between sheep populations in the Iron Age eastern Mediterranean.”

Additionally, it would be beneficial to delve into hypotheses and factors explaining the obtained results. Consider discussing anthropogenic and environmental factors influencing phenotypic variations in livestock.

The subject of phenotypic variation is now addressed by a short exposition with eight additional references at the beginning of the introduction. Further down the same section, we added sentences explaining why the astragali from the sites we have chosen to focus on are suitable for an analysis of inter-population disparity, while referring again to various factors affecting growth and form. However, we do not envision this paper as a review of these topics, which are well and amply explained in the references we cite.

The introduction contains valuable insights but requires reorganization. Consider placing objectives after establishing the context for better flow. Avoid incorporating methodological details; focus on the significance of geometric morphometrics without delving into procedural descriptions. It would be relevant to mention Ariadna Nieto Espinet's work on Iron Age livestock movements in Catalonia and Languedoc.

We prefer to state the objectives of the research very briefly, as we do here, already at the beginning of the manuscript. However, the research objective is re-iterated towards the end of the introduction. The explanation on the calculation of Procrustes disparity was deleted and added to the methods section. Nieto-Espinet's work in Languedoc is now cited in the first paragraph of the introduction.

Lines 17-18: The link between the scarcity of studies on Mediterranean fauna and the maritime connections in the Eastern Mediterranean is missing in the abstract/context. Clarify the relationship between these two aspects.

The connection was emphasized by rephrasing the first two sentences of the abstract, which now read:

"The movement of livestock across the Mediterranean is well-documented in the Neolithic era, but its significance during subsequent periods has received less attention. Here we start addressing this lacuna by analyzing astragal bone morphology from four coastal and inland sites in Israel and Cyprus, seeking potential evidence for maritime connections between sheep populations in the Iron Age eastern Mediterranean."

Line 23: It would have been interesting to elaborate further on the hypotheses and the different factors that may explain the results you found.

The factors affecting shape and size, of phenotypic variation, are all potential hypotheses to explain variability we may observe in astragular form. The factors are now listed at the beginning of the introduction. Further down the same section, we added sentences explaining why the astragali from the sites we have chosen to focus on are suitable for an analysis of inter-population differences in shape that are not related to nutrition, sexual dimorphism, age, etc.,

while referring again to various factors affecting growth and form. These are not phrased as official null hypotheses, but rather as controlled (through choice of element and samples) sources of variation.

Line 25: Perhaps it could be interesting to discuss the different factors involved in the phenotypic variations of livestock (anthropogenic and/or environmental).

See immediately above.

Lines 26-28: The statement is possibly true for the Eastern Mediterranean but seems less applicable to the Western Mediterranean (eg the work of Silvia Valenzuela, Ariadna Nieto Espinet, Allowen Evin, Angela Trentacoste…).

Definitely!! We changed “traditionally” to “until recently”, to clarify that interest in post-Neolithic livestock populations’ comparative biometry and morphology has peaked in the last decade, mostly due to the work of the archaeozoologists mentioned, who are cited throughout the introduction (and also pioneers like Davis and Albarella).

Lines 29-31: The sentence is a bit complicated, containing a lot of information. It might be preferable to split it into several sentences.

The sentence has been split as requested: “Mediterranean maritime trade in the Bronze and Iron Ages immediately evokes the riches of the Uluburun shipwreck (Cline & Yasur-Landau, 2007; Pulak, 1998). However, smaller-scale trade is known from the late second millennium BCE onward in the eastern Mediterranean by evidence of the mixed cargoes of amphorae, scrap metal and other mundane items found, for example, in the Cape Gelidonya and Point Iria shipwrecks (Bass, 1967; Bass, 1961; Phelps et al., 1999).”

Lines 36-39: The objectives come a bit early, without a clear context beforehand. Inverting the order might improve fluidity. The context should precede the objectives. Line 51: It might be interesting to mention the work of Ariadna Nieto Espinet on livestock movements during the Iron Age in Catalonia and Languedoc.

We prefer to state the objectives of the research very briefly, as we do here, already at the beginning of the manuscript. However, the research objective is re-iterated towards the end of the introduction. Ariadna’s work is cited in the first paragraph of the introduction. Our context is maritime and eastern Mediterranean, so we did not go into details that might depart too much from that ballpark.

Line 75: Be careful, morphotype and phenotype are not synonymous. Phenotype is a broader term encompassing all observable traits, including morphological traits. Morphotype specifically refers to morphological traits.

“Morphological/phenotypic” was changed for “morphological”, to avoid ambiguity.

Line 75: In my opinion, it would be more relevant to present the objectives at this stage.

Indeed, the objective is mentioned just below: “In this study, we directly address the relationship between morphological variation in livestock and access to maritime trading networks in antiquity.”. We prefer, however, to retain the brief reference to the objective in the first paragraph of the introduction.

Line 75: Avoid citing figures in the introduction. This figure relates solely to the material and methods section. Be careful to make the distinction (likewise for Lines 79-82). Overall, the distinction between the introduction and the material and methods in this paragraph is not clear. Stick to context for the introduction, while this paragraph already delves into methodological considerations.

Citation of figure 1 was moved to the “Sites” section. The methodological comments were removed.

Lines 82-90: Be cautious about shortcuts involving size= human impact & shape=phylogeny. An object's shape can also be influenced by other factors (e.g., pathology in bovines, mobility, etc.). GMM simply separates these two components, providing distinct information. Lines 92-99: Methodological considerations.

The shortcuts that the reviewer mentioned were removed.

Lines 101-118: We haven't seen the results yet, or any chronology. This will be discussed later in the discussion

Moved to the discussion.

Materials & Methods

Globally, the material and methods section would benefit from more clarity. It is challenging to navigate, and some parts are missing, particularly the section on statistics and graphical representations. Methodological considerations are crucial for a clear understanding of the presented results.

Lines 121-123: Introductory contextual elements.

Moved to the introduction

Line 132: Add a reference to Figure 1.

Added

Line 136: "with refs." What does it mean?

Reference manager output, removed.

Figure 1: It is easy to confuse sea/land. Why not use clearer colors? (e.g., light blue for the sea, brown for the land, for example). You mention different cities; they could be added to the map for a clearer interpretation.

We prefer the minimalist (and very standard) template from Wikimedia. Given the proximity of our sites in Israel to each other on this scale, two additional colors for sea and land would clutter the figure — as would modern cities that would cluster with the ancient sites.

Line 167: Why only one site from Cyprus? Are there no other sites that could be included in these analyses?

Not at hand, but we will try for future studies.

Line 226: Why choose the first 10 axes and not another number (arbitrary choice)?

It is an arbitrary choice, which captures 85% of the variation.

Line 228: What tests are performed to answer your hypothesis? We have no data on the statistics used. They appear gradually with the results, while we should already know in the methods what will be tested and how. We have no information on the visualizations that will be used (CVA? PCA? Why?).

We think that the statistical tests and explanation of the visualization methods are succinctly explained in the methods section, cited below. They all rely on standard procedures in GMM, which we do not think require lengthy explanation.

"Generalized Procrustes analysis. A Generalized Procrustes Analysis (GPA) was performed using the gpagen() function in 'geomorph' (Adams et al., 2022; Adams & Otárola-Castillo, 2013) on the landmark coordinates obtained in tpsDig which produced Procrustes shape variables (Bookstein, 1991). Semi-sliding landmarks were allowed to slide relative to each other during the GPA in order to minimize the sum Procrustes distances between each specimen and the mean shape (Bookstein, 1996). Centroid sizes of the specimens were also produced by the same function. The effect of size, group structure, and the interaction between them on the shape coordinates was estimated using the residual randomization permutation procedure ANOVA method using the geomorph::procDlm function.
Ordination and disparity analyses. We used principal component analysis (PCA, implemented using 'geomorph::gm.prcmp') to reduce the dimensionality of the Procrustes transformed landmarks, while retaining most of the variability present among the Procrustes shapes. The first ten principle components were used in a canonical variates analysis ('Morpho::CVA') to

visualize group structure and measure inter-group Mahalanobis distances in the resulting ordination space. We also conducted a Procrustes disparity analysis on the Procrustes transformed coordinates using geomorph::morphol.disparity. Morphological disparity is estimated as the Procrustes variance for groups, using residuals of a linear model fit. P-values can be obtained using a random residual permutation procedure. In itself, Procrustes variance is the trace of the group covariance matrix divided by the number of observations in the group (Zelditch et al., 2014, p. 487).”

Results

In general, for the results, it is challenging to draw meaningful conclusions without a clear understanding of the statistical methods. The section combines methodological aspects (presentation of performed statistics), result presentation, and comparisons with other studies, which should be reserved for the discussion. Therefore, this section needs refinement to clearly delineate its purpose: presenting the results only. A clearer presentation of the methods and statistics used will enhance the readability of the results.

Line 234: The calculation of the error percentage is introduced in the results, whereas it should be explained in the methods. Consider using Claude's method (CLAUDE, J. 2008. Morphometrics with R. Springer-Verlag New York. Montpellier: Springer-Verlag New York.). Additionally, a nearly 30% variation due to landmark placement seems substantial. How does this compare to studies using the same protocol? Have you conducted tests with different individuals?

The calculation of the percentage error is now fully explained in the methods: “Digitization percentage errors were checked before statistical analysis by using tpsSmall64 (v.1.0) and tpsrelw32 (v.1.53) (Rohlf, 2015) according to the protocol described by Hulme-Beaman (2014: 153). A subset (N=28) of specimens from the dataset were digitized three unique times to test for digitization error; these were randomly selected but as evenly distributed between the study sites as possible. This analysis employs Procrustes distances to quantify both biological variation and digitization error. Firstly, it calculates the Procrustes distance between all specimens to assess the extent of inherent biological variability. Secondly, it computes the Procrustes distance between replicate measurements of each individual, isolating the variability introduced by the digitization process. Finally, by expressing the digitization error as a percentage of the total observed variability, the impact of measurement imprecision on the overall results is evaluated.”

We did not find mention of measurement error analysis in Claude’s 2008 textbook, but elsewhere he uses ANOVA on centroid sizes to determine the relative amount of variability explained by within- and among-individual components. This is not essentially different from Hulme-Beaman’s protocol that we used. The 27% variability due to measurement error is not a far cry from that reported in a previous study by SH on late antique ship rats (*Rattus rattus*) (Harding, 2021, p. 59) or by Hulme-Beaman in a population of *R. exulans* (22%; Hulme-Beaman

2014: table 6.2) using the same protocol. Hulme-Beaman has also noted that the percentage measurement error is also a function of the number of landmarks (as opposed to semi-sliding landmarks) in the digitization protocol. The amount of variation in our shape data that is not explained by the significant components of size and group affinity are very large (84%), and the measurement error probably accounts for a hefty portion of it.

The respective part of the results now reads: “The mean Procrustes distance between re-digitized specimens was 0.034, and among the full dataset, 0.118; the percent of digitization error is therefore estimated at 29.37%. This percent of digitization error is within the range observed in similar intraspecies GMM studies (Harding, 2021, p. 59; Hulme-Beaman, 2014, pp. 164–167 and table 6.2).”

Lines 242-248: These sentences involve interpretation and should be moved to the discussion section. Results should only contain a description of the obtained results without comparisons with other studies, which should be reserved for the discussion.

The sentences were moved to the discussion.

Line 248: It is unclear what you are testing or how, as it has not been explained in the methods. Therefore, it's uncertain whether such statements can be made.

Now clarified in the Methods section: “The effect of size, group structure, and the interaction between them on the shape coordinates was estimated using the residual randomization permutation procedure ANOVA method using the geomorph::procDlm function.”

Figure 3: It could be beneficial to use different symbols for different types of sites (coastal or inland) for easier and quicker interpretation. Adding significant differences in pairwise tests between sites could also enhance the figure.

Another level of graphical coding will create visual clutter. The simplified structure can be seen in the dendrogram. We added significant differences as requested, and amended the caption accordingly.

Lines 259-260: The information about the geographical structure similarity between CVA and the neighbor-joining tree has already been mentioned in the preceding paragraph. This similarity isn't surprising, given that both representations are based on discriminant analysis. These are two different representations that can complement each other.

The neighbor-joining tree in Figure 4A is not based on the CVA values, but on the Procrustes distances between groups (using the geomorph::morphol.disparity function). It reflects a similar structure because it is inherent in the data, as our analysis shows. Displaying this tree here is also handy as a strut for displaying the astragular morphologies.

Line 261: Throughout the text, you refer to your GMM points as landmarks. However, according to Pollath et al.'s protocol, there are also sliding semi-landmarks, which have different characteristics than landmarks. Be cautious about this quick shortcut.

We add the following sentence to the methods section: “Digitization of landmarks and semi-sliding landmarks (collectively ‘landmarks’ below, for brevity) was conducted by laboratory assistant Daria Loshkin Gorzilov, under the direct supervision of SH, using tpsDig232 (v.231) (J. F. Rohlf, 2017).”

Figure 4A: It is unclear why there is another tree compared to Figure 3C (although the visualization of Figure 4 is preferred). What new information does this figure provide? However, an interesting aspect is that inland sites are much closer to each other than coastal sites.

The neighbor-joining tree in Figure 4A is not based on the CVA values, but on the Procrustes distances between groups (using the geomorph::morphol.disparity function). It reflects a similar structure because it is inherent in the data, as our analysis shows. Displaying this tree here is also handy as a strut for displaying the relationship between astragal morphologies. The seeming redundancy also helps the reader by having an immediate and simplified visual reference to the data structure within each figure.

Table 2: Some numbers are in italics. Are these significant figures? If so, some are missing. How are the pairwise absolute differences between sample variances calculated? What are the differences with p-values, and why include both?

The italicized values above the diagonal are significant p-values. Although we do not see any significant values that were not italicized, we removed the italics to avoid the apparent confusion, and disambiguated the relation to the diagonal in the caption.

Discussion

In my opinion, the discussion lacks references for authors to discuss their results, compare them to lend weight, and contextualize them within previous research. It also seems challenging to address the initially posed question in this paper. Additionally, it might be interesting to include potential avenues for future research.

Lines 291-296: The authors' intended message is not entirely clear. It might be worthwhile to refine the bibliography in this regard.

The sentence was corrected, and now reads: “First, with respect to size, we observe a significant difference between sites, with the Cypriot LTD showing notably smaller size than the other, southern Levantine sites. The astragalus geometric morphometric results suggest that slightly larger sheep were present in Tell Keisan, which is like the result obtained by a biometric

study of the sheep remains from the site (Harding et al., 2023). The smaller size of the Dor specimens also resonates well with the results from other Levantine coastal sites in the Iron Age (Chahoud et al., 2023)." To the best of our knowledge, there have not been large scale studies of sheep morphometry in the Levant. The way for that opened recently, with the launching of the ZooLog R package developed by Valenzuela-Lamas and colleagues. Both cited papers are from a special issue devoted to its application.

Line 295: Does this trend of greater variation on coastal sites exist elsewhere? For other periods? The result is presented here but not thoroughly discussed.

No, to the best of our knowledge. We highlighted it and the sentence now reads: "Second, morphological variability appears to be greater in coastal sites compared to inland sites, suggesting a morphological cline that to the best of our knowledge was not noted before."

Line 294: Indeed, the site with the largest sample comes from a single ensemble, so it seems normal for the variance to be relatively low.

Indeed, we discuss this below and devote a figure to the explanation.

Lines 295-298: Similarly, the result is not discussed or compared to other bibliographic references, which could add weight to the presented results.

While the exact pattern noted here does not have immediate parallels, we added five references that demonstrate the prevalence of morphological clines through space and time in archaeozoology.

Lines 300-305: This paragraph presents important elements but lacks references.

Four references added.

Figure 5: This figure has not been presented in the results nor mentioned. It appears only in the discussion. A presentation of the method in the appropriate section, followed by the presentation of results, seems more fitting. Also, there is uncertainty about understanding this figure.

The figure relates to a discussion of the results, and hence its place in the preprint. It illustrates how variability is related to archaeological context. We clarified the caption, which now reads: "The effect of zooarchaeological context on intra-site shape variability when sample sizes are similar. Rectangle areas, representing intra-site morphological variability, are a function of the spatial variability (x-axis) of the region from which the animals were obtained (large for the ritual site of LTD) and the time period (y-axis) over which they accumulated (very short in ABM)."

Supplementary Data

In the sheep_specinfo, linear measurements such as GLl, Bd, etc., are recorded, but it appears they are not utilized in the paper. Could you provide insights into what additional information these measurements could have offered alongside GMM? Alternatively, why include them in the supplementary data if they are not analyzed?

We would agree if the specimen info was presented in the main article. However, this is a supplement, and including these linear measurements — basic to the description of bones in archaeozoological databases — does not harm, and might even help someone in the future.

During geometric morphometrics analysis, be attentive to removing duplicate points. For instance, in the sliding procedure, landmarks 1 & 3 are duplicated (refer to attached figure). To address this, remove slidings 12 & 25 to avoid having points counted twice, which could introduce bias.

1 and 11 are part of the constellation of fixed landmarks, while 12 to 25 are the start and end points of a curve. We anchored it on purpose at the fixed landmarks 1 and 3, otherwise we would have a severe digitization bias. Although we are not sure why a consistent placing of landmarks the way we have done could introduce any bias, we “unslided” those points to check and found that the differences would be almost imperceptible: please compare the resulting ANOVA tables when “unsliding” the landmarks (top) vis a vis the original.

Because we're not sure that we agree with the need to unslide the points, we would rather retain the procedure that we have used.

```
             Df      SS         MS     Rsq      F      Z Pr(>F)
group         3 0.06485 0.0216154 0.11458 5.9279 6.1116  0.001 **
Csize         1 0.00765 0.0076498 0.01352 2.0979 2.1232  0.014 *
group:Csize   3 0.01579 0.0052630 0.02790 1.4433 1.4924  0.066 .
Residuals   131 0.47768 0.0036464 0.84401
Total       138 0.56596
            Df      SS         MS     Rsq      F      Z Pr(>F)
group         3 0.06398 0.0213280 0.11534 5.9762 5.9044  0.001 **
Csize         1 0.00781 0.0078066 0.01407 2.1874 2.2188  0.009 **
group:Csize   3 0.01541 0.0051373 0.02778 1.4395 1.4720  0.068 .
Residuals   131 0.46752 0.0035688 0.84280
Total       138 0.55472
---
```

It would be beneficial to present the results of ANOVA in a table, either within the results section or as supplementary data, with clear and visible formatting rather than coded (with a legend).

Of course! An ANOVA table appears in the supplementaries, in addition to the output to the R-studio console when running the code. The formatted table includes details of the analysis and codes for the levels of significance.

```
Analysis of Variance, using Residual Randomization
Permutation procedure: Randomization of null model residuals
Number of permutations: 1000
Estimation method: Ordinary Least Squares
Sums of Squares and Cross-products: Type I
Effect sizes (Z) based on F distributions

             Df      SS        MS     Rsq      F      Z Pr(>F)
group         3 0.06398 0.0213280 0.11534 5.9762 5.9044  0.001 **
Csize         1 0.00781 0.0078066 0.01407 2.1874 2.2188  0.009 **
group:Csize   3 0.01541 0.0051373 0.02778 1.4395 1.4720  0.068 .
Residuals   131 0.46752 0.0035688 0.84280
Total       138 0.55472
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Call: procD.lm(f1 = coords ~ group * Csize, RRPP = TRUE, data = model_df)
```

Additionally, the ANOVA indicates differences between groups, but have you investigated whether these differences are due to inter-site variations regardless of the group? Running pairwise tests between each site could provide more detailed insights.

Agreed. The third page of the supplementary materials indeed provides such post-hoc significance tests (Tukey’s HSD):

File: size_ANOVA.txt

```
                 Df Sum Sq Mean Sq F value Pr(>F)
spec_info$group   3  12.64   4.213   46.63 <2e-16 ***
Residuals       135  12.20   0.090
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = gpa_out$Csize ~ spec_info$group)

$`spec_info$group`
                      diff         lwr        upr     p adj
Cyprus-ABM    -0.79339407 -0.98729098 -0.5994971 0.0000000
Dor-ABM        0.06490874 -0.16025097  0.2900685 0.8765657
Keisan-ABM     0.21904806  0.01658192  0.4215142 0.0284066
Dor-Cyprus     0.85830281  0.58583570  1.1307699 0.0000000
Keisan-Cyprus  1.01244213  0.75840676  1.2664775 0.0000000
Keisan-Dor     0.15413932 -0.12449099  0.4327696 0.4773647
```

In the code, you calculate disparity in two different ways. The first is based on the linear model, while the second relies on Procrustes coordinates. However, you only present results based on Procrustes coordinates. Could you elaborate on why both methods are included?

We assume that the reviewer refers to the part of the script pasted below. The answer is that the analysis of disparity based on procrustes coordinates takes as input the object (“mod_1”) created by the geomorph::procD.lm function, which is the linear model relating coordinates to size, group, and their interaction. Both results are presented: the linear model yielded the ANOVA results explained in the text and tabulated in the supplementary (see above); and the procrustes distances appear in the neighbor-joining tree (figure 4A) and Table 2.

```
#Shape ANOVA
 print("calculating Procrustes distance linear model in geomorph, shape ~ group * centroid size...")
 model_df <- geomorph.data.frame(gpa_out, group = spec_info$group)
 mod_1 <<- procD.lm(coords ~ group * Csize, data = model_df, RRPP = TRUE)
 print(summary(mod_1))
 print("created 'procLM.txt with the results.")

 setwd(output_dir)
 sink(file = "procLM.txt")
 print(summary(mod_1))
 sink()
```

```
setwd(working_dir)

#Disparity
print("calculating morphological disparity with geomorph, call: shape ~ group * size")
disparity_1 <<- morphol.disparity(mod_1, groups = model_df$group)
print(disparity_1)
plot(nj(dist(disparity_1$PV.dist)), "u")
print("calculating morphological disparity with geomorph, call: shape ~ group")
disparity_2 <<- morphol.disparity(gpa_out$coords ~ 1, groups = model_df$group)
print(disparity_2)
plot(nj(dist(disparity_2$PV.dist)), "u")
setwd(output_dir)
```

Adding PCA as supplementary data would enhance the completeness of the analysis. The supplementary data is not well-explained. Providing clear annotations and legends for this data would improve its clarity and interpretation.

PCA biplot added to the supplement, and is outputted by the script for those interested in more details. Captions added.

Reviewer #2

The paper submitted by Harding et al. "The contribution of Mediterranean connectivity to morphological variability in Iron Age sheep of the Eastern Mediterranean", addresses the question of the movement of flocks by sea around the Mediterranean during the Iron Age. It examines this question by analyzing the morphology of sheep talus from two coastal sites and two inland sites in Israel and Cyprus. The authors hypothesize that the intra-site morphological variability of this bone must be more significant on the coastal sites than the inland sites due to the greater genetic diversity of sheep populations on the coast exposed to commercial networks. To highlight this variability, the authors use a geometric morphometric approach, which should better analyze the phylogenetic signal by isolating the variables of shape and size. Their results support the initial hypothesis, as the data show more significant variability at coastal sites.

The authors present the issues related to the studied assemblages and remain cautious in the interpretation of the results. Except for one site (ABM), the number of sites and bones per site is minimal, and the chronology is very broad, covering several centuries with the pooling of phases.
The study supports the idea that maritime connectivity is related to morphological variability in the Iron Age. Later studies should verify these results by analyzing more Iron Age II assemblages.
Despite the limited number of sheep talus and archaeological sites, this study merits attention. It explores a new approach using GMM to investigate the role of domestic animals in trade and commerce in the Mediterranean during the Iron Age, providing promising results.

Please, note that the following references cited in the text are missing from the bibliography:

Aubet, 2014
Bartosiewicz and Lisk, 2018
Briend and Humbert, 1980
Gambash, 2015
Gilboa, 2015
Gilboa et al., 2015
Gilboa and Sharon, 2017
Lehmann, 2001; 2021
Lehmann and Peilstocker, 2012
Nitschke et al., 2011
Panitz-Cohen et al., 2013;
Panitz-Cohen and Yahalom-Mack, 2019
Raban, 1981
Raban and Galili, 1985
Raban-Gerstel et al., 2008
Raveh and Kingsley, 1991
J. F. Rohlf, 2017

Sapir-Hen et al., 2014
Stern et al., 1993a
Stern et al., 1993b
Wachsmann and Raveh, 1984
Yahalom-Mack et al., 2018
Yasur-Landau et al., 2018

Thank you for your thoughtful assessment of our paper and pointing out the missing references from the bibliography. They have now been added and the text double-checked for reference accuracy.

Reviewer #3

This paper discuses the morphological variability of sheep astralali bones from Levantine sites to test the hypothesis that coastal sites will have a higher varibility of livestock introduced by trade. The authors use GMM to undertake detailed analysis of shape of the bones and, despite the small sample sizes and equifinality of the underlying causes of variation, this paper discusses is a novel application of the method. The paper is primilinary as noted by the authors but will be a useful starting point for more detailed analyses with larger samples sizes, better dated material and a comparison with aDNA data in the future will make results more meaningful. However, the paper is a very useful publication as a starting point for more detailed analysis.

Thank you for your supportive comments on our manuscript. We also hope to see further future studies with larger sample sizes and complementary methods such as palaeogenomics, etc.

Reviewer #4

Preprint strengths:

• The issue of the impact of maritime trade on the spread of livestock and the evolution of domestic sheep breeds is very interesting and innovative from the point of view of geometric morphometrics methods.
• Chronological period poorly documented for these lines of research
• Fluid writing, easy to read

Preprint weaknesses:
• Problems with the structure of the article
• Insufficient justification for statistical tests
• Socio-cultural heterogeneity of the sites selected and, in fact, of astragali corpus
• The discussion section lacks important data, such as the impact of the taxonomic identification of the astragali, or the links between the heterogeneity of the socio-cultural contexts of the corpus and the results obtained in GMM, even though an effort has been made to highlight this.

GENERAL COMMENTS:

Title
No comments.

Abstract
No comments.

Introduction, Materials & Methods:
The article presents a structural problem in the "study sites", "material" and "methods" sections. Many sentences describing the corpus and the method are in the introductive part. You could move the methodological sections written in paragraphs 5 and 6 of the introductions to the method section to present GMM and justify its use.

Accepted, this part of the manuscript has changed following the comments by Reviewer #1 (above).

Similarly, the sites presentation should be moved to the material section, where table 1 referring to sampling should be referred to in the text (data currently missing).

Site descriptions were moved into the materials and methods section, and we made reference in the text to Table 1.

I suggest the following structure: 1. Present the corpus, adding information on the temporal resolution of the data + the origin of the assemblages (paragraph 2 of the material and method section). 2. Present the GMM method + statistical tests and justify it.

The structure was changed as requested.

2 Regarding the methodological presentation, in my mind it lacks many justifications for the choice of statistical tests used and bibliographical references. I didn't find any references on GMM. These need to be added.

Reviewer #1 also requested numerous clarifications regarding the statistical tests and methods, which we incorporated as described above. We are surprised that references to GMM were not found. We cited both archaeozoological works (and more were added following Review #1) and basic texts that appeared in the original version of the preprint (see below). It is important to note that we did not apply any technique which is off the beaten track (i.e. is not implemented and documented in a widely-used R library).

1. Bookstein, F. L. (1991). Morphometric Tools for Landmark Data . Cambridge University Press.
2. Bookstein, Fred L. (1996). Landmark methods for forms without landmarks: localizing group differences in outline shape. Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis , 279–289.
3. Cardini, A., Seetah, K., & Barker, G. (2015). How many specimens do I need? Sampling error in geometric morphometrics: testing the sensitivity of means and variances in simple randomized selection experiments. Zoomorphology , 134 (2), 149–163.
4. Rohlf, J. (2015). The tps series of software. Hystrix , 26 (1), 9–12.
5. Zelditch, M. L., Swiderski, D. L., & Sheets, H. D. (2014). Geometric Morphometrics for Biologists: A Primer (2nd ed.). Academic Press.

A methodological presentation of the various statistical tests that are used in the results section to discuss the data are missing: centroid size, ANOVA, permutation tests, allometry tests, visualisation of unrooted morphological proximity trees.

Added in the methods section:

Centroid size – under “generalized procrustes analysis”; ANOVA + permutation tests + allometry – under “Procrustes analysis of variance (ANOVA)”; nj trees under “ ordination and disparity analyses”.

Similarly, the method used to produce the average conformation patterns per site and their visualisations is not explained.

It is the mean shape obtained from the output of the geomorph::gpagen function. Now explained in the methods under “generalized procrustes analysis”

The sites are not presented in the same terms. In introduction view, I recommend that you provide additional information describing the topographical characteristics of the environment in order to justify your choice of your samples when you write « we chose samples that derive from

sites in a similar topographic setting ». I would also advise you to add information on the main findings of the archaeozoological studies to give an overview of animal economy identified.

The site descriptions were homogenised so that they now read at the same level of details. General observations on distance from the sea, terrain, and vegetation cover were added. A sentence summary of the zooarchaeological findings was added.

Furthermore, if it is possible, I advise you to add as co-authors of the publication your colleagues who participated in the acquisition of the morphometric data on which your results are based.

We added Mr. Roee Shafir, the lab technician at the university of Haifa who photographed all the specimens (including in Cyprus).

Discussion: Overall, interpretations need to be nuanced due to corpus heterogeneity (bones from totally different contexts). Even though these arguments are set out in the third paragraph, I suggest that you detail your hypotheses on morphotypical ovine diversity, taking these limitations into account.

We have added a paragraph detailing this topic: “An additional factor that may affect ovine astragalar morphotypical diversity is the landscape to which domestic sheep may be adapted to over generations, i.e., functional mobility adapted to a rough, hilly terrain (closed environment) versus flat plains (open environment) may be reflected in astragalar morphology (Haruda et al., 2019). In the present study, ABM would be considered the most topographically complex, being a high-elevation settlement in the Galilee mountain range (414 m above mean sea level [AMSL]). Dor is located on a coastal promontory (3 m AMSL), and Keisan is situated in a valley near the Galilee foothills (31 m AMSL). Although coastal in terms of linear distance, LTD sits 85 m AMSL in a hilly area of the coastal plain. If functional mobility was a strong factor affecting astragalar morphology, we would expect to see that signal clearly reflected in the data, e.g., no significant difference between ABM and LTD or between Dor and Keisan. Our results show a significant pairwise difference between ABM and LTD ($p=0.001$), but not between Dor and Keisan. A previous study echoes this finding, which analyzed astragalar proportions of inland/rough terrain and coastal/flat terrain domestic sheep in the southern Levant using an astragalar dimension index (lateral depth/greatest lateral length; distal breadth/lateral depth) (Harding et al., 2023). The

proportions of astragali tended to reflect the environment from which the samples came, yet there was significant overlap in the center of the distribution which muddied any stark distinction between open and closed landscapes (Harding et al., 2023: 8, Fig. 5). Based on the current findings and this previous study, we do not find sufficient evidence to support the hypothesis that terrain-adapted functional morphology overwhelmingly accounts for the variability that we observe in the present study. "

I would also advise you to write some details of the interpretative problems associated with taxonomic identification that might suggest the variability of the data within sites. Although it is doubtful that the outliers from the LTD site were wild animals, this hypothesis cannot be refuted due to the anatomical identifications.

A short note to that effect was added to the "Specimens" section of the methods.

The text now reads "The possibility that wild sheep were included in the analysis is very small for the sites in Israel, where no wild sheep existed during the Iron Age. No mouflon horncores have been identified during the ongoing analysis of LTD faunal remains by AH. A spatiotemporally comparable assemblage currently under study in Cyprus by AH consists of only domestic animals which tend to be young in age, further indicating the preference for domestic rather than feral sheep. While some specimens from LTD could represent mouflon (feral sheep), given the absence of any game animal from the site, we find it unlikely."

Figure 5: explain in the text.

Done, in the discussion.

DETAILS COMMENTS:
Introduction:
« to minimize possible functional morphological differences » = justify

This paragraph has been re-written after Review #1, and the sentence no longer appears in the text.

« Our results, which represent the first study of animal mobility in the southern Levant using geometric morphometric methods, should be somewhat liberally interpreted due to the primacy of this study »: I think the term 'liberally interpreted' is completely inappropriate. On the contrary,

as the data set is heterogeneous and limited, and as no comparative study has yet been carried out on the influence of maritime mobility on the introduction of new sheep varieties thanks to GMM, we need to be extremely cautious about the results.

This part of the introduction was edited out following Review #1. We agree with the critique, the term “liberal” is out of place in this context.

3 Methods:
« osteologically mature at time of death » = to be justified due to the age identification problems based on talus ossification.

A few words on the young age of ossification was added.

« Ovis sp. » = explain why the identification remained at the gender level and why this may be problematic in relation to the hypotheses formulated in the article.

We assume “genus”. The possibility of mouflons is now brought up in the next sentence: “The possibility that wild sheep were included in the analysis is very small for the sites in Israel, where no wild sheep existed during the Iron Age. No mouflon horncores have been identified during the ongoing analysis of LTD faunal remains by AH. A spatiotemporally comparable assemblage currently under study in Cyprus by AH consists of only domestic animals which tend to be young in age, further indicating the preference for domestic rather than feral sheep. While some specimens from LTD could represent mouflon (feral sheep), given the absence of any game animal from the site, we find it unlikely.”

Better justify the use of CVA.

The use of CVA is ubiquitous for visualization of group structure in gmm analyses. We nevertheless add a clause to the methods:

The first ten principle components were used in a canonical variates analysis (‘Morpho::CVA’) – a standard technique that maximizes the distance between groups – to visualize group structure and measure inter-group Mahalanobis distances in the resulting ordination space.

Can you include the results of the PCA in the additional data?

Yes, done (see Review #1)

Results:
Overall, in this section, you should remove the comparisons with other studies and describe the results objectively.

Done, see also Review #1

Digitization error
Can you also indicate the result obtained on landmarks alone (without taking sliding into account).

The issue of digitization error has been referred to extensively in our reply to Review #1. Our standard protocol (Pollath, 2019) uses semi-sliding landmarks, so we see little point in measuring the error without them.

Centroid size « The smaller size of the Dor specimens also resonates well with the results from other Levantine coastal sites in the Iron Age (Chahoud et al., 2023). The specimens from Cyprus are especially small. We do not know how universal this pattern may be on the island during the Iron Age, and this observation requires further investigation » = for discussion

Done, following Review #1

« The effect of the interaction between group and size on shape is not significant » = indicate the pvalue.

Added (P = 0.068)

You should make the results of the allometric tests more obvious. I recommend adding this information to the subtitle « centroid size » and “allometric pattern”.

Done.

Ordination: In view of the CVA results, I doubt the need to add the tree calculated from the Mahalanobis distances from the CVA. I advise you to add this graph and this information to the "supplementary" section.

We agree that the dendrogram is formally redundant. However, displaying the same information in different ways helps figure readability and highlights the pattern of differences between sites. We hope that adding the P-values alleviated this issue to some extent.

« The group-based CVA ordination of the first ten principal components explains almost all the variability in that dataset » pay attention to the formula. The first two axes of the CVA performed on the first 10 dimensions of the PCA explain 98% of the variability.

Indeed, rephrased: “The first two canonical variates capture most of the variability in the dimensionality-reduced dataset of the first 10 PCs (~98%)(Figure 3B).”

Disparity: Be careful with the formulation: here the data are calculated on the basis of Procrustes distances using which method (PCA)?

No, they are calculated from the landmark configurations (technically, by the Procrustes ANOVA procedure). Clarified: “The Procrustes distances between groups, calculated from the landmark configurations, […]”

4 « The similarity between the results of two different analyses—direct measurements of disparity on landmark data and on eigen-ordinated coordinates—support the statistical results »: add the raw data.

References were inserted to Table 2 and Figure 3. The confusion was caused because Table 2 was not cited in the text.

Figure 3: I advise you to divide up this figure for greater clarity and visibility. You could, for example, present the boxplots and then present the conformation data in a second figure. In addition, I would suggest that you remove the individual numbers from the CVA biplot and, if possible, use colour plates to bring the groups together.

We prefer not to accept these suggestions. The figure clearly summarizes the size and shape results at the level of the group. The numbers in the CVA plot are necessary for those who wish to see the location of specific astragali in relation to the spec_info file, and are not savvy enough to modify the R script. We have used a color palate from the ‘ggsci’ package to the plot.

Figure 4B is not explained in the text: either explain it and therefore add the method that enabled it to be carried out in the method section, or remove it. In my own view, this description is very important in order to show the differences in shape that exist between the 'average' sheep astragalus at each of the sites.

It was added to the methods section: “The mean Procrustes shape for each site was obtained from the geomorph::gpagen function for visualization.” And explained in the caption.

REFERENCES

All the references mentioned are appropriate and accurate. However, GMM references are missing. You have to add some references concerning the origin of the method like (Rohlf and Marcus, 1993, Bookstein, 1991…) If you want, you can add too the PhD thesis of Manon Vuillien next to the reference of Colominas et al., 2019 and Haruda et al., 2019 : https://www.theses.fr/2020COAZ2020 She works on the morphological variability of Late Neolithic and Iron Age sheep in Provence combining GMM and more traditional approaches in archaeozoology.

Thank you for bringing that Manon’s dissertation to our attention. We have a reviewed it and added it to the reference list. We have also ensured that Bookstein 1991 and Zelditch 2014 are listed in the bibliography.